Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 08–12

ON THE CHOICE OF THE SMOOTHING PARAMETER FOR THE BHEP GOODNESS-OF-FIT TEST

CARLOS TENREIRO

Abstract: The BHEP test for assessing univariate and multivariate normality introduced by Epps and Pulley (Biometrika 70 (1983) 723) and Baringhaus and Henze (Metrika 35 (1988) 339) has shown to be a relevant test procedure, recommended in some recent comparative studies. It is well known that the finite sample behaviour of the BHEP goodness-of-fit test strongly depends on the choice of a smoothing parameter h. In this paper we give a theoretical and finite sample based description of the role played by the smoothing parameter in the detection of departures from the null hypothesis of normality. Additionally, we develop a Monte Carlo study in order to propose an easy to use rule for the choice of h that produces a test with the omnibus property. In the important multivariate case, and contrary to the usual choice of h, the BHEP test with the smoothing parameter proposed in this paper presents a comparative good performance against a wide range of alternative distributions. In practice, if no relevant information about the tail of the alternatives is available, the use of this new bandwidth is strongly recommended. Otherwise, new choices of h which are suitable for short tailed and long tailed alternative distributions are also proposed.

Keywords: BHEP goodness-of-fit test, kernel density estimator, Bahadur eficiency, multivariate normality, Monte Carlo power comparison.
AMS Subject Classification (2000): 62G10, 62G20.

1. Introduction

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed d-dimensional random vectors with unknown density function f. Following an idea of Anderson, Hall and Titterington [1] that have used kernel density estimators with fixed bandwidth for testing the equality of two multivariate probability density functions, Fan [17] uses the Bickel-Rosenblatt statistic (cf. Bickel and Rosenblatt [10]; see also [16, 20]) with a constant bandwidth for testing the composite hypothesis that f is a member of a general parametric family of density functions. The test statistic is based on the L_2 distance between the kernel density estimator and an estimator of its

Received January 29, 2008.
This work was partially supported by CMUC (Centre for Mathematics, University of Coimbra)/FCT

mathematical expectation under the null hypothesis and it is given by

$$I_n^2(h) = n \int \{f_n(x) - \bar{\mathrm{E}}_0 f_n(x)\}^2 dx, \tag{1}$$

where the unspecified integral denotes integration over the whole space,

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i),$$

for $x \in \mathbb{R}^d$, $K_h(\cdot) = K(\cdot/h)/h^d$ with K a kernel, that is, a bounded and integrable function on $\mathbb{R}^d$, and h is a strictly positive real number.

A location-scale invariant version of the previous statistic was first considered by Bowman and Foster [11] (with $h \to 0$) to test a multivariate normality hypothesis. For a location-scale family $g(\cdot; m, S)$, and denoting by

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$$

the sample mean vector and by

$$S_n = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X}_n)(X_j - \bar{X}_n)',$$

the sample covariance matrix where the prime denotes transpose, the location-scale invariant version of $I_n^2(h)$, called IBRF statistic, is given by

$$\mathbf{I}_n^2(h) = n|S_n^{1/2}| \int \{\bar{f}_n(x) - \varphi_n(x; \bar{X}_n, S_n)\}^2 dx, \tag{2}$$

where, for $x \in \mathbb{R}^d$,

$$\bar{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{H_n}(x - X_i),$$
$$K_{H_n}(\cdot) = |H_n^{-1}| K(H_n^{-1} \cdot),$$
$$H_n = h\, S_n^{1/2},$$

with $h > 0$, $S_n^{1/2}$ is the square root of S_n and, for $m \in \mathbb{R}^d$ and S a nonsingular covariance matrix,

$$\varphi_n(x; m, S) = \int K_{H_n}(x - y) g(y; m, S) dy.$$

For a general location-scale family $g(\cdot; m, S)$ and a general kernel K, Tenreiro [35] gives the limiting null distribution of $\mathbf{I}_n^2(h)$, the consistency of the associated tests and derive its asymptotic power against sequences of local alternatives.

In this paper our attention will be focused on the important case where the previous parametric family is the Gaussian:

$$g(\cdot; m, S) = |S^{-1/2}|\, \phi(S^{-1/2}(\cdot - m)), \tag{3}$$

where ϕ is the d-variate Gaussian standard density

$$\phi(x) = (2\pi)^{-d/2} \exp(-\tfrac{1}{2}x'x),$$

for $x \in \mathbb{R}^d$. Although an extensive literature exists regarding the testing of normality, there is a continued interest in this classical problem as attested by the recent papers of Mecklin and Mundfrom [28], Arcones [2], Arcones and Wang [3], Farrel, Salibian-Barrera and Naczk [19], Arcones [4], Coin [12] and Yazici and Yolacan [36].

Concerning the IBRF statistic several simple choices for K are possible. The choice $K = \phi$ is particularly interesting when $g(\cdot; m, S)$ is the Gaussian family because in that case the calculation of $\mathbf{I}_n^2(h)$ does not require any integration (cf. [11, 24]):

$$\mathbf{I}_n^2(h) = \frac{1}{n} \sum_{i,j=1}^{n} \mathbf{Q}(Y_i, Y_j; h), \tag{4}$$

with

$$Y_i = S_n^{-1/2}(X_i - \bar{X}_n),$$

and

$$\mathbf{Q}(u, v; h) = \phi_{(2h^2)^{1/2}}(u - v) - \phi_{(1+2h^2)^{1/2}}(u) - \phi_{(1+2h^2)^{1/2}}(v) + \phi_{(2+2h^2)^{1/2}}(0),$$

for $u, v \in \mathbb{R}^d$.

In this form the statistic $\mathbf{I}_n^2(h)$ have been previously considered by several authors and is known in the literature as BHEP statistic where the smoothing parameter is usually denoted by β and is connected with h through the relation $h = 1/(\beta\sqrt{2})$ (see Henze [24]). It has been introduced in the real case by Epps and Pulley [15] and in the multivariate case by Baringhaus and Henze [8] and Henze and Zirkler [23] as an empirical characteristic function based goodness-of-fit test. In fact, similarly to what was pointed out by Fan [17] in relation to statistic (1), the general IBRF statistic (2) can be interpreted as a L_2 weighted distance between the empirical characteristic function and the parametric estimate of the characteristic function implied by the null model with weight function $t \to |\hat{K}(th)|^2$, where $\hat{K}$ is the Fourier transform of K. Theoretical properties about the BHEP test are given in Baringhaus and Henze [8], Csörgő [13], Henze and Zirkler [23], Henze [24] and Henze and Wagner [25].

As noticed in some of previous references, the empirical power of the test based on $\mathbf{I}_n^2(h)$ varies considerably with the parameter h. In the univariate case, Epps and Pulley [15] considered the values $h = 0.35, 0.49, 0.71, 0.92$ and conclude that with $h = 0.49$ or 0.71 the test is an omnibus normality test, whereas for $h = 0.35$ and 0.92 the test appears to be suitable for platykurtic and for leptokurtic distributions, respectively. In the multivariate case Henze and Zirkler [23] conclude that for $h = 1.41$ the test is powerful against heavy tailed distributions, but they also report some extremely poor results for some alternatives where a small choice for h is much better. However, after the initial papers about the BHEP test, it became a standard procedure to consider in practice the values $h = h_{\mathrm{EP}} := 0.71$ and $h = h_{\mathrm{HZ}} := 1.41$ for the univariate and multivariate cases, respectively, as we can see in the papers of Baringhaus *et al.* [9], Arcones and Wang [2], Mecklin and Mundfrom [28] and Farrel *et al.* [19].

The main purpose of this paper is to examine these usual choices of the smoothing parameter h and, if necessary, to propose alternative choices for h that produce a test that is generally powerful against a wide range of alternatives. This question is particularly relevant in the multivariate case where the BHEP has shown to be a relevant test procedure, recommended in some recent comparative studies like those of Mecklin and Mundfrom [28] and Farrel *et al.* [19].

Although less interesting in practice we start our study by the analysis of the univariate case that we present in Section 3. This case will give us some important guide lines about the test behaviour as a function of the smoothing parameter h that will have natural counterparts in a multivariate context. Contrary to the above mentioned studies about the choice of h that are exclusively based on Monte Carlo experiments, we begin our study by computing the Bahadur approximate indices of the univariate BHEP statistic for a set of Edgeworth alternatives. For that we use a result presented in Section 2 and proved in Section 5 about the Bahadur approximate slopes of the general IBRF statistic. These indices enables us to get a better understanding on the role played by the smoothing parameter in the detection of departures from the null hypothesis of normality in terms of each one of the third, fourth, fifth and sixth moments. They suggest that a large, but not too large, bandwidth is adequate for detection of departures from normality in skewness and kurtosis, whereas a small, but not too small, bandwidth is adequate for detection of hight moments alternatives (see also Tenreiro [34] for the test of a simple hypothesis of normality). These conclusions agree with the finite sample power properties of the BHEP test that indicate that a small bandwidth should be used for short tailed alternatives whereas a large bandwidth is suitable for long tailed or skewed ones.

Taking into account the previous results, a Monte Carlo experiment based on a set of univariate alternative distributions which are members of the generalized lambda family, was conducted to propose a practical choice of the bandwidth. A similar simulation study is undertaken in Section 4 for the multivariate case ($2 \leq d \leq 15$). Here we have considered a set of meta-Gaussian distributions whose marginal distributions are the univariate lambda distributions taken in the univariate case analysis.

Two distinct behaviour patterns were observed for the BHEP empirical power as a function of h which lead us to propose two distinct choices of the bandwidth, depending on the data dimension, which are suitable for short tailed and long tailed alternative distributions. If we do not have relevant information about the tail of the alternative distributions, we propose to consider the mean value of the two previous bandwidths, $\bar{h}$, as rule-of-thumb for the choice of h. This way we intend to obtain a test with a reasonable performance against both types of alternatives.

In the univariate case, our proposal is identical to the one by Epps and Pulley [15]. In the multidimensional case $\bar{h}$ behaves much better than h_{HZ} for short tailed alternatives and is only slightly inferior than h_{HZ} for long tailed alternatives. In practice, we strongly recommend the use of $\bar{h}$ that should replace, in a multivariate context, the usual choice of the bandwidth in the BHEP goodness-of-fit test. With this new bandwidth the BHEP test revealed a comparative good performance against a wide range of alternative distributions.

2. Bahadur efficiency

In order to compare the test based on $\mathbf{I}_n^2(h)$ given by (2) with other test procedures, or to compare $\mathbf{I}_n^2(h)$ tests obtained for different values of h, we give in this section the Bahadur approximate slope $C_{\mathbf{I}(h)}$ of the so-called standard sequence of tests $\mathbf{I}(h) = (\mathbf{I}_n(h))$, where $\mathbf{I}_n(h) = \mathbf{I}_n^2(h)^{1/2}$. Given two different tests, the one with highest Bahadur approximate slope is preferred. We restrict our attention to the case of the Gaussian family (3) but the following result is also valid for a general location-scale parametric family. For the description of Bahadur's concept of efficiency, see Bahadur [5, 6, 7] or Nikitin [29].

Let us denote by $\mathcal{K}$ the set of bounded and integrable functions K on $\mathbb{R}^d$ with continuous partial derivatives up to order three, such that the functions defined for $u = (u_1, \dots, u_d)'$, by

$$u \to \sup_{H \in V_I} \|u\|^k \, |\frac{\partial^k K}{\partial u_{i_1} \dots \partial u_{i_k}}(Hu)|,$$

for $k = 1, \dots, \ell$ and $i_1, \dots, i_k = 1, \dots, d$, where $\| \cdot \|$ is a norm in $\mathbb{R}^d$ and V_I is some neighbourhood of the identity matrix I, are bounded and integrable on $\mathbb{R}^d$. Remark that every three times continuously differentiable function with compact support belongs to $\mathcal{K}$. The standard normal density function also belongs to $\mathcal{K}$. More generally, every spherical function $K(x) = k(x'x)$, where $k : \mathbb{R}^+ \to \mathbb{R}$ is three times continuously differentiable and each derivative $\psi^{(\ell)}$ is nonincreasing and satisfies $\int (x'x)^\ell \, |k^{(\ell)}(x'x)| dx < \infty$, for $\ell = 0, 1, 2, 3$, belongs to $\mathcal{K}$.

Denote by $\mathcal{D}$ the set of all bounded probability density functions f on $\mathbb{R}^d$ such that $\int \|x\|^2 f(x) \, dx < +\infty$ with nonsingular covariance matrix and by $\mathcal{D}_0$ the subset of $\mathcal{D}$ of all densities that take the form (3) for some $m \in \mathbb{R}^d$ and some nonsingular covariance matrix S. For $f \in \mathcal{D}$ let us write f^* the density in $\mathcal{D}$ defined by $f^*(x) = |S_f^{1/2}| f(m_f + S_f^{1/2} x)$, for $x \in \mathbb{R}^d$, where m_f and S_f are the mean and covariance matrix of f.

The next result follows from Gregory's [21] and Bahadur's [5] results. Its proof is given in Section 5.

Theorem 1. *Let $K \in \mathcal{K}$ with Fourier transform $\hat{K}$ such that the set $\{t \in \mathbb{R}^d : \hat{K}(t) = 0\}$ has Lebesgue measure zero. For $f \in \mathcal{D} \setminus \mathcal{D}_0$ we have*

$$C_{\mathbf{I}(h)}(f) = \lambda_{1,h}^{-1} \, b_{\mathbf{I}(h)}(f),$$

where

$$\begin{aligned} b_{\mathbf{I}(h)}(f) &= \int |K_h \star f^*(x) - K_h \star \phi(x)|^2 dx \\ &= (2\pi)^d \int |\hat{f}^*(t) - \hat{\phi}(t)|^2 |\hat{K}(ht)|^2 dt, \end{aligned}$$

and $\lambda_{1,h}$ is the largest eigenvalue of the symmetric positive semidefinite Hilbert-Schmidt operator A_h defined, for $q \in L_2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Phi) =: L_2(\Phi)$, by

$$(A_h q)(u) = \int \mathcal{Q}(u, v; h) q(v) d\Phi(v),$$

where $\Phi = \phi\lambda$ with λ the Lebesgue measure in $\mathcal{B}(\mathbb{R}^d)$ and $\mathcal{Q}(\cdot, \cdot; \cdot)$ is the function defined, for $u, v \in \mathbb{R}^d$ and $h > 0$, by

$$\mathcal{Q}(u, v; h) = \int \mathcal{K}(x, u; h)\mathcal{K}(x, v; h) dx, \tag{5}$$

where for $u = (u_1, \ldots, u_d), x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $h > 0$,

$$\begin{aligned} \mathcal{K}(x, u; h) &= K_h(x - u) - K_h \star \phi(x) + \sum_{p=1}^{d} K_h \star \phi_p(x)\, u_p \\ &\quad + \frac{1}{2} \sum_{1 \le p \le q \le d} K_h \star \phi_{pq}(x)\, (u_p u_q - \delta_{pq}), \end{aligned}$$

with $\phi_p(x) = -x_p \phi(x)$, $\phi_{pq}(x) = (x_p x_q - \delta_{pq})\phi(x)$ and δ_{pq} the Kronecker symbol.

If the null density ϕ belongs to a family of probability density functions of the form $\{f(\cdot; \theta) : \theta \in \Theta\}$, where Θ is a nontrivial closed real interval containing the origin and $\phi = f(\cdot; 0)$, it is natural to compare the set of competitor tests $\mathbf{I}^2(h), h > 0$, through its Bahadur local approximate slopes $C_{\mathbf{I}(h)}(f(\cdot; \theta))$ when $\theta \to 0$. Under some regularity conditions on the previous parametric family, two useful representations can be given for $C_{\mathbf{I}(h)}(f(\cdot; \theta))$.

Assumptions on $\{f(\cdot; \theta) : \theta \in \Theta\}$ (P)

For all $\theta \in \Theta$, $f(\cdot; \theta)$ has zero mean and identity covariance matrix, and for all $x \in \mathbb{R}^d$ the function $\theta \to f(x; \theta)$ is continuously differentiable on Θ, and there exists a neighbourhood of the origin $V \subset \Theta$ such that the function $x \to \sup_{\theta \in V} |\frac{\partial f}{\partial \theta}(x; \theta)|$ is integrable on $\mathbb{R}^d$.

Under assumption (P), we have

$$C_{\mathbf{I}(h)}(f(\cdot; \theta)) = \lambda_{1,h}^{-1} \int \left| K_h \star \frac{\partial f}{\partial \theta}(\cdot; 0)(x) \right|^2 dx\, \theta^2 (1 + o(1)), \theta \to 0. \tag{6}$$

We conclude that the Bahadur local approximate slopes take the form $\theta^2(1 + o(1))$, up to the multiplication by a constant, when $\theta \to 0$, and therefore for the comparison of tests $\mathbf{I}^2(h)$ it is sufficient to compare the coefficients of θ^2 called *approximate local indices.*

An informative alternative representation can be given for the previous local approximate slope. Denote by $\{q_{k,h}, k \in \mathbb{N}_0\}$ the orthonormal basis of $L_2(\Phi)$ consisting of eigenfunctions of A_h, i.e., for all k and j, $\int \mathcal{Q}(\cdot, v; h) q_{k,h}(v)\, d\Phi(v) = \lambda_{k,h} q_{k,h}$, *a.e.* (Φ) and $\langle\, q_{k,h}, q_{j,h}\,\rangle = \delta_{kj}$, where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $L_2(\Phi)$. Under assumption (P), if $\frac{\partial \ln f}{\partial \theta}(\cdot; 0) \in L_2(\Phi)$ the approximate local indice $C_{\mathbf{I}(h)}(f(\cdot;\theta))$ can be expressed in terms of the weights $(\lambda_{k,h})$ and the principal components $(q_{k,h})$:

$$C_{\mathbf{I}(h)}(f(\cdot;\theta)) = \sum_{k=1}^{\infty} \lambda_{1,h}^{-1}\, \lambda_{k,h}\; a_{k,h}^2\, \theta^2 (1 + o(1)), \theta \to 0, \tag{7}$$

where

$$a_{k,h} = \Big\langle\, q_{k,h}, \frac{\partial \ln f}{\partial \theta}(\cdot; 0)\, \Big\rangle,$$

for $k = 1, 2, \ldots$.

If the eigenvalues sequence $(\lambda_{k,h})$ converges quickly to zero, it is clear from this representation that only a finite number of alternatives directions effectively contribute to $C_{\mathbf{I}(h)}(f(\cdot;\theta))$. As we will see in the next section, in the particular case of the BHEP statistic (4), the weights $\gamma_{k,h} = \lambda_{1,h}^{-1}\, \lambda_{k,h}$, for $k = 2, 3, \ldots$, strongly depend on the choice of h.

3. Testing a univariate hypothesis of normality

In this section, in the particular case of testing a univariate hypothesis of normality, we use the results of the previous section to get a better understanding of the role played by the smoothing parameter in the detection of departures from the null hypothesis. From now on we take for K the standard normal density $K = \phi$. Therefore $\mathbf{I}_n^2(h)$ is the BHEP statistic (4).

3.1. Most significant weights. As described in Section 2, the Bahadur approximate slopes of $\mathbf{I}^2(h)$ depend on the weights $(\gamma_{k,h})$, where $\gamma_{k,h} = \lambda_{1,h}^{-1}\, \lambda_{k,h}$, and on the principal components $(q_{k,h})$. Numerical evaluations of the most significant weights are shown in Table 1 for four values of h. These approximations have been obtained along the lines described in Tenreiro [34]. For comparison and following Stephens [33] we also give the most significant weights for the Anderson-Darling test of normality.

	$h = 0.01$	$h = 0.1$	$h = 1.0$	$h = 2.0$	A^2
$\gamma_{2,h}$	9.99×10^{-1}	9.77×10^{-1}	4.54×10^{-1}	1.75×10^{-1}	7.33×10^{-1}
$\gamma_{3,h}$	9.71×10^{-1}	7.34×10^{-1}	5.31×10^{-2}	7.20×10^{-3}	3.65×10^{-1}
$\gamma_{4,h}$	9.70×10^{-1}	7.11×10^{-1}	2.08×10^{-2}	1.09×10^{-3}	2.95×10^{-1}
$\gamma_{5,h}$	9.44×10^{-1}	5.43×10^{-1}	3.31×10^{-3}	6.31×10^{-5}	1.84×10^{-1}
$\gamma_{6,h}$	9.42×10^{-1}	5.22×10^{-1}	1.20×10^{-3}	8.75×10^{-6}	1.61×10^{-1}
$\gamma_{7,h}$	9.17×10^{-1}	4.03×10^{-1}	2.18×10^{-4}	5.91×10^{-7}	1.17×10^{-1}
$\gamma_{8,h}$	9.15×10^{-1}	3.86×10^{-1}	7.55×10^{-5}	7.74×10^{-8}	1.02×10^{-1}
$\gamma_{9,h}$	8.91×10^{-1}	3.00×10^{-1}	1.48×10^{-5}	5.69×10^{-9}	7.90×10^{-2}
$\gamma_{10,h}$	8.89×10^{-1}	2.86×10^{-1}	4.95×10^{-6}	7.15×10^{-10}	7.04×10^{-2}
$\gamma_{11,h}$	8.65×10^{-1}	2.24×10^{-1}	1.03×10^{-6}	5.50×10^{-11}	5.70×10^{-2}
$\gamma_{12,h}$	8.64×10^{-1}	2.12×10^{-1}	3.33×10^{-7}	6.57×10^{-12}	5.16×10^{-2}

TABLE 1. Weights $\gamma_{k,h}$ for $\mathbf{I}(h)$

From these values and the representation (7) for the Bahadur local approximate slopes, we expect that the test based on $\mathbf{I}_n^2(h)$ for small values of h could use information contained in others components different from the first ones. This conclusion agrees with the asymptotic properties of the test based on the invariant Bickel-Rosenblatt statistic $\mathbf{I}_n^2(h_n)$ where h_n tends to zero when n goes to infinity (see [10, 35]). For moderate or large values of h, $\mathbf{I}_n^2(h)$ might exclusively use information contained in the first components.

3.2. Approximate local indices. Using representation (6), in the following we compute the approximate local indices of $\mathbf{I}^2(h)$ for a set of local alternatives which are based on the Edgeworth series for the density and express departures from the null hypothesis in the jth moment (about Edgeworth expansion see Hall [22] and the references therein). We assume that the local alternatives satisfy (P) with $f = f_j$, and are such that

$$(A.\,j) \qquad \frac{\partial \ln f_j}{\partial \theta}(\cdot;0) = H_j(\cdot)/\sqrt{j!}\,, \tag{8}$$

for $j = 3, \dots, 6$, where H_j is the jth Hermite polynomial defined by

$$\begin{aligned} H_3(x) &= x^3 - 3x; \\ H_4(x) &= x^4 - 6x^2 + 3; \\ H_5(x) &= x^5 - 10x^3 + 15x; \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15. \end{aligned}$$

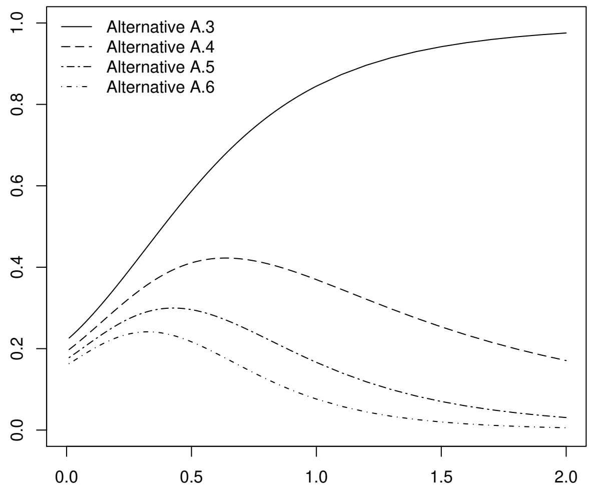

FIGURE 1. Approximate local indices for $\mathbf{I}(h)$, $h \in [0.01, 2]$

The skeweness and kurtosis alternatives considered by Durbin *et al.* [14] satisfy $(A.3)$ and $(A.4)$, respectively.

The approximate local indices plotted in Figure 1, for $h \in [0.01, 2]$, agree with the previous conclusions. It is clear that a large, but not to large, bandwidth is adequate for detection of departures from normality in skewness and kurtosis, whereas a small, but not too small, bandwidth is adequate for detection of hight moment alternatives.

3.3. The finite sample performance. The main purposes of this section are to know if the finite sample power performance of the $\mathbf{I}^2(h)$ tests for fixed alternatives is in accordance with the theoretical properties based on Bahadur efficiency and also to provide some suggestions on the choice of h.

In order to examine the performance of the $\mathbf{I}^2(h)$ tests for alternatives in skewness and kurtosis, we consider four symmetric and four asymmetric alternative distributions (four of them are platykurtic and the other four are leptokurtic), whose densities are shown in Figure 2. These distributions are members of the generalized lambda family discussed in Ramberg and Schmeiser [31]. The distributions of this family are easily generated because they are defined in terms of the inverses of the cumulative distribution functions: $F^{-1}(u) = \lambda_1 + (u^{\lambda_3} - (1-u)^{\lambda_4})/\lambda_2$, for $0 < u < 1$. The parameters defining the distributions used in the study and the associated moments μ_i of orders three to six are given in Table 2. All the considered distributions have zero mean and unit variance. Some of these distributions are used in

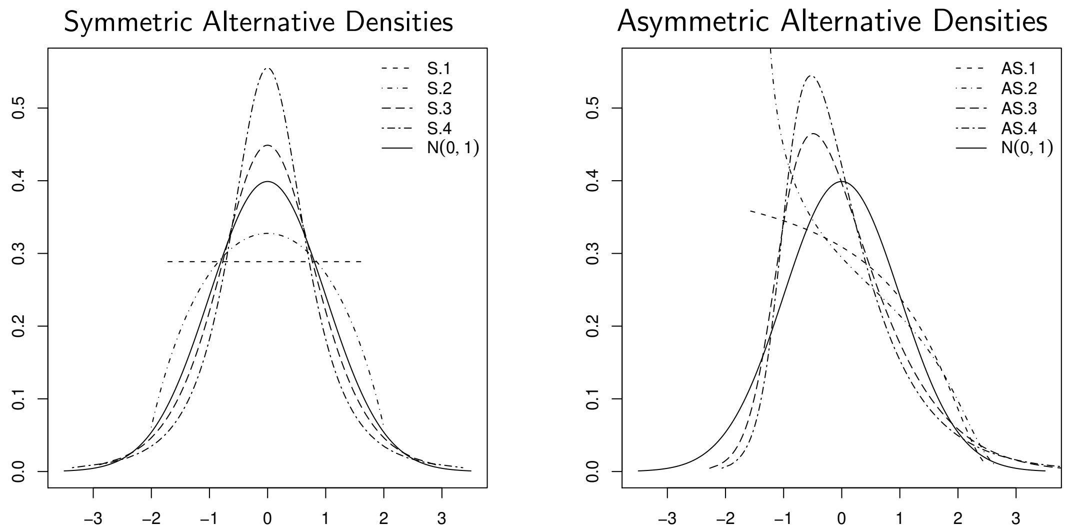

FIGURE 2. Densities used in the simulation study

Fan [16] to examine the performance of the Bickel-Rosenblatt test with a bandwidth converging to zero as n tends to infinity.

In Figure 3 we present the empirical power of $\mathbf{I}_n^2(h)$ as a function of h with $h = 0.05, (0.05), 2$, at level 0.05, for some of the previous distributions for $n = 40, 60, 80, 100$. For the evaluation of the critical values of $\mathbf{I}_n^2(h)$ we have used 10^4 repetitions generated by the R function rnorm (cf. [30]) and the Monte Carlo power results are based on 2000 samples from the considered set of alternatives.

Case	μ_3	μ_4	μ_5	μ_6	λ_1	λ_2	λ_3	λ_4
	Symmetric distributions							
S.1	0	1.80	0	3.86	0	0.577350	1	1
S.2	0	2.08	0	5.54	0	0.463251	0.5	0.5
S.3	0	4.06	0	35.90	0	0.017829	0.01	0.01
S.4	0	11.61	0	7211.95	0	−0.397012	−0.16	−0.16
	Asymmetric distributions							
AS.1	0.31	2.09	1.69	6.05	0.578020	0.465781	1	0.3
AS.2	0.51	2.22	2.74	7.40	0.835034	0.459063	1.4	0.25
AS.3	0.90	4.21	10.32	40.88	−0.635145	0.096880	0.025	0.094
AS.4	1.52	7.46	33.83	209.04	−0.628997	−0.037156	−0.0075	−0.03

TABLE 2. Distributions used in the simulation study

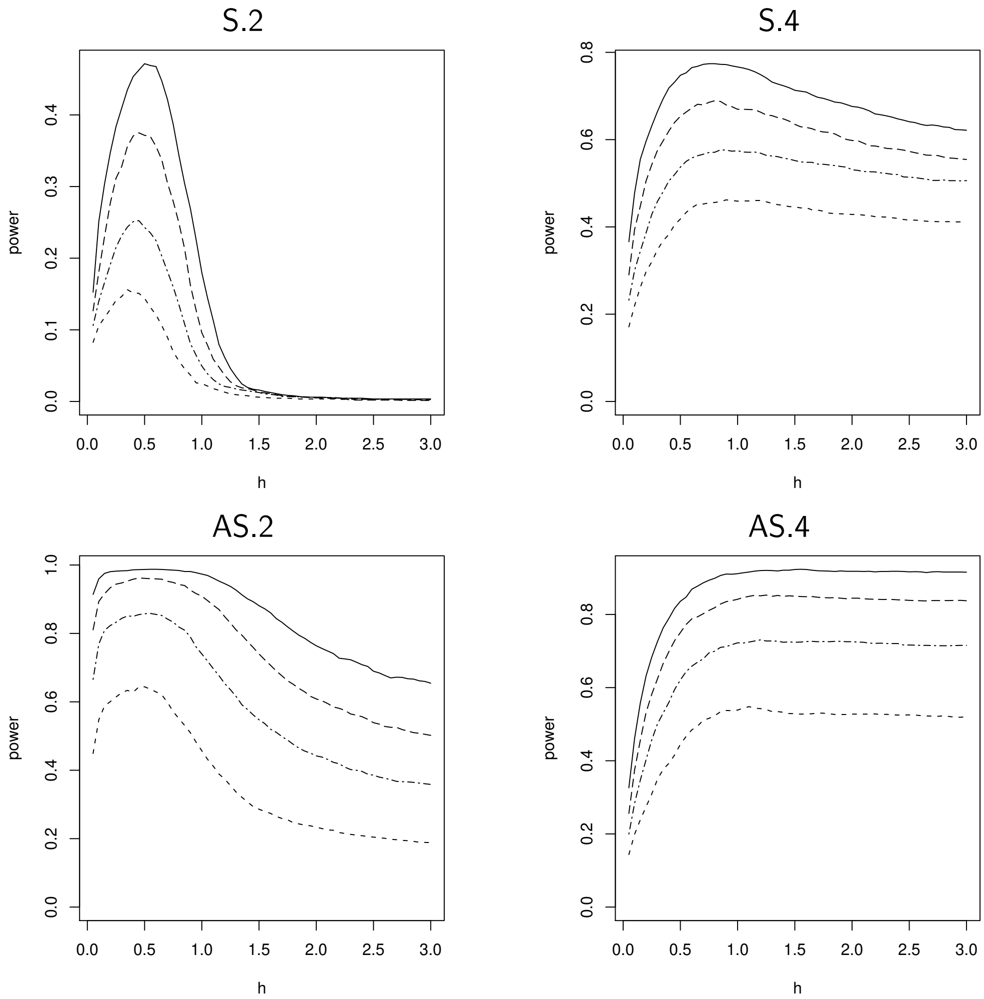

FIGURE 3. Empirical power as a function of h for $n = 100$ (solid line), $n = 80$ (large broken line), $n = 60$ (broken and dotted line), $n = 40$ (short broken line), at level 0.05.

These empirical results are globally in accordance with the theoretical ones based on Bahadur efficiency. For short tailed alternatives their departure from normality in skewness and kurtosis is necessarily small which explains the hight power observed for small (but not too small) values of h and the low power observed for large values of h. This is particularly true for symmetric alternatives. For long tailed, symmetric or asymmetric, alternatives a large value for h should be chosen. As mentioned in the introductory section, these conclusions are also supported by the simulation studies undertaken by Epps and Pulley [15] and Baringhaus *et al.* [9].

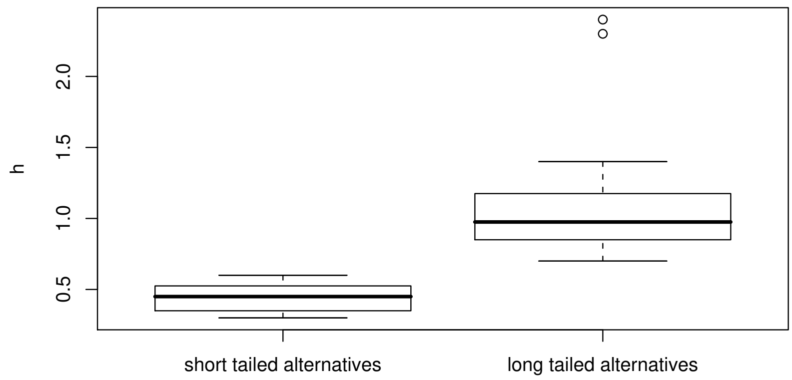

FIGURE 4. Distribution of $h_{f,n}$ at level 0.05

3.4. A rule-of-thumb for the choice of h. In order to get an easy-to-use rule for choosing the bandwidth h, for each one of the previous alternative distributions and for sample sizes $n = 20, 40, 60, 80, 100$, we compute the nearly optimal bandwidth $h_{f,n}$. This bandwidth is obtained from the bandwidth $h^*_{f,n}$ that maximizes the empirical power over the set $H = \{0.05, 0.1, 0.15, \dots, 3.0\}$ of values of h by taking $h_{f,n} = h^*_{f,n}$ for short tailed alternatives and $h_{f,n} = \inf\{h \in H : p(h) > 0.995\, p(h^*_{f,n})\}$, where $p(h)$ denotes the empirical power of the BHEP test associated to the bandwidth h, for long tailed alternatives. This way we avoid very large bandwidths that do not lead to a significant power improvement (see distribution AS.4 in Figure 3).

We remark that $h^*_{f,n}$ does not depend significantly on the sample size n but it strongly depends on the underlying alternative distribution tail type. The sample distribution of the bandwidths $h_{f,n}$ is described in Figure 4 for the BHEP test at level 0.05 for short tailed and long tailed alternatives, respectively.

As a rule-of-thumb for the choice of h for these to types of underlying distributions we propose to use the medians of the previous empirical distributions. For short tailed distributions we obtain $h_{\mathrm{S}} = 0.45$, whereas for long tailed distributions we get $h_{\mathrm{L}} = 0.975$. If we do not have relevante information about the alterative distribution, we propose to use for smoothing parameter the mean value of the two previous bandwidths $\bar{h} = (h_{\mathrm{S}} + h_{\mathrm{L}})/2 \approx 0.71 = h_{\mathrm{EP}}$. This proposal leads to the choice of h suggested by Epps and Pulley [15] and also considered in other studies like those of Baringhaus *et al.* [9] and Arcones and Wang [3].

3.5. Finite sample power analysis. Based on a simulation study undertaken for a large number of univariate alternative distributions usually considered in power studies for testing univariate normality, we conclude that the choices h_{S} and h_{L} present good performance for short tailed and high moment alternatives and for long tailed alternatives, respectively. Moreover, $\bar{h}$ present a reasonable performance against both types of alternatives. These conclusions agree quite well with the results described in Epps and Pulley [15].

We limit ourself to present in Figure 5 the empirical power results obtained for some of the considered alternatives. They give us a good picture about the overall observed power results. The uniform and beta distributions are short tailed distributions whereas the student and lognormal distributions are long tailed one (symmetric and asymmetric ones, respectively). The considered generalized exponential power distributions, GEP, whose distribution shape depends on two parameters, are high moment alternatives. They have zero mean, unit variance and their third and fourth order moments are equal to the ones of the normal distribution (see Johnson *et al.* [26] about the GEP distribution family).

The evaluation of the critical values of $\mathbf{I}_n^2(h)$ for $h = h_{\mathrm{S}}, h_{\mathrm{L}}, \bar{h}$ was based on 10^5 repetitions under the null hypothesis and the power results are based on 10^4 samples of different sizes from the considered set of alternatives. With 10^4 repetitions, the margin of error for approximate 95% confidence intervals for the proportion of rejections do not exceed 0.01.

Finally we remark that the previous rules-of-thumb can also be used for the BHEP test at levels 0.01 and 0.1, since the distributions of the bandwidths $h_{f,n}$ for these two levels are very similar to that one shown in Figure 4.

4. Testing a multivariate hypothesis of normality

The test of a multivariate normality hypothesis is discussed in this section. Contrary to the real univariate case, where theoretical and simulation based results have been used to study the test properties, the performance of the multivariate BHEP test is assessed exclusively through a Monte Carlo study.

However, it is natural to expect that the behaviour of the test $\mathbf{I}_n^2(h)$ as a function of h described in the univariate case could have a natural counterpart in a multivariate context. A first indication in this direction is given by the simulation study undertaken by Henze and Zirkler [23] who considered the bandwidths $h = 0.23, 0.71, 1.41$ and compared these tests with other

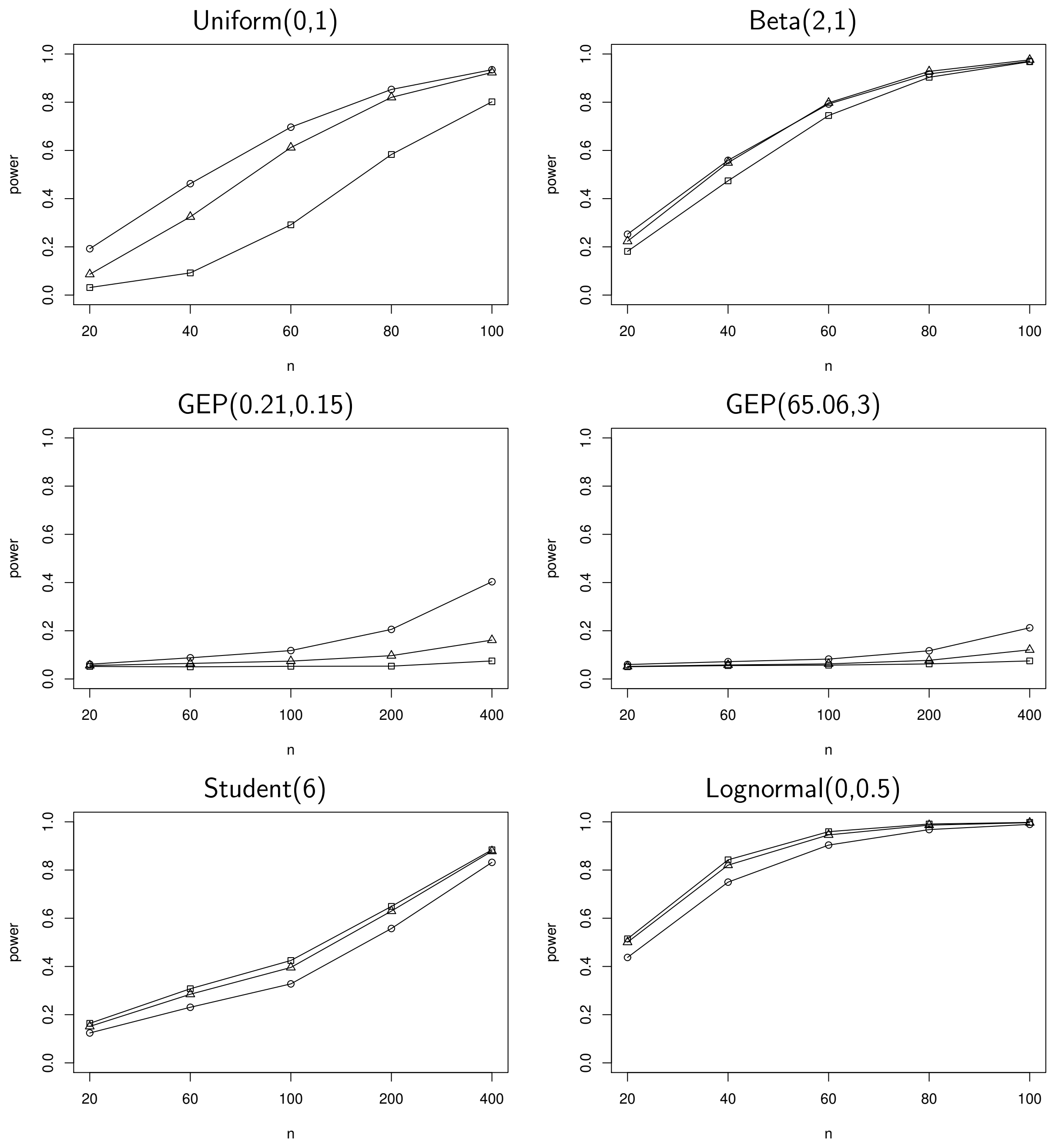

FIGURE 5. Empirical power for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$ and $h = \bar{h}(\triangle)$

procedures for assessing multivariate normality. The test with $h = 0.23$ was generally inferior to the other two BHEP tests and the choice $h = 1.41$ gives excellent results for asymmetric or long tailed distributions but also very weak results for symmetric and short tailed distributions.

4.1. A rule-of-thumb for the choice of h. In order to propose a practical rule for the choice of the bandwidth, we conduct a simulation study based on a set of meta-Gaussian distributions whose marginal distributions are the univariate distributions considered in paragraph 3.3 (see Fang *et al.* [18] for the idea of meta-type distribution). If $Z = (Z_1, \ldots, Z_d)$ is a multivariate normal distribution with mean zero and unitary diagonal covariance matrix Σ, and F is the cumulative distribution function of one of the generalized lambda distributions S.1,...,AS.4, the alternative distributions we consider are the distributions of the vector $X = (X_1, \ldots, X_d)$, defined by $X_i = F^{-1}(\Phi(Z_i))$, for $i = 1, \ldots, d$, where Φ is the cumulative distribution function of the univariate standard normal distribution. Since we consider $\Sigma = [\sigma_{ij}] = \Sigma_\rho$ such that $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho$, with $0 \leq \rho < 1$, for $1 \leq i \leq j \leq d$, the previous distributions will be denoted by S.$1_d(\rho)$,...,AS.$4_d(\rho)$.

After some simulation work it appears that the value of h that maximizes the empirical power of the BHEP test does not depend significantly on the sample size n. However, it depends on the data dimension, and mainly on the underlying alternative distribution. Moreover, as in the univariate case, if we want a test with the omnibus property, extreme choices of h have to be avoided. These facts are illustrated in Figure 6 where we present the two typical behaviours, for short tailed and long tailed distributions, of the BHEP empirical power as a function of h, for $n = 40, 60, 80, 100$, $d = 2, 5, 10$ and $h = 0.05, (0.05), 3.0$. For the evaluation of the critical values we generated 10^4 samples using the R function mvrnorm (cf. [30]), and the power estimates are based on 2000 samples from the considered set of alternatives.

In general, the bandwidth that maximizes the power seems to be an increasing function of the data dimension d. However, it is interesting to remark the large empirical power obtained for small values of h for the short tailed distributions S.$1_d(\rho)$, S.$2_d(\rho)$,AS.$1_d(\rho)$,AS.$2_d(\rho)$ when ρ is small and d is large. This type of behaviour was for the first time pointed out by Henze and Wagner [25] for $d = 5$ for the uniform distribution over the unit d-cube and for a symmetric Pearson Type II distribution. A nonasymptotic argument can explain this type of behaviour. For a fixed n, from representation (4) we see that for large values of d and depending on the sparseness of the observations, the term $n^{-1} \sum_{1 \leq i < j \leq n} \phi_{(2h^2)^{1/2}}(Y_i - Y_j)$ loose its influence and the sum $\sum_{i=1}^n \phi_{(1+2h^2)^{1/2}}(Y_i)$ determines the random behaviour of test statistic $\mathbf{I}_n^2(h)$ for small values of h. Assuming a normal approximation for the distribution of previous sum, we conclude that the

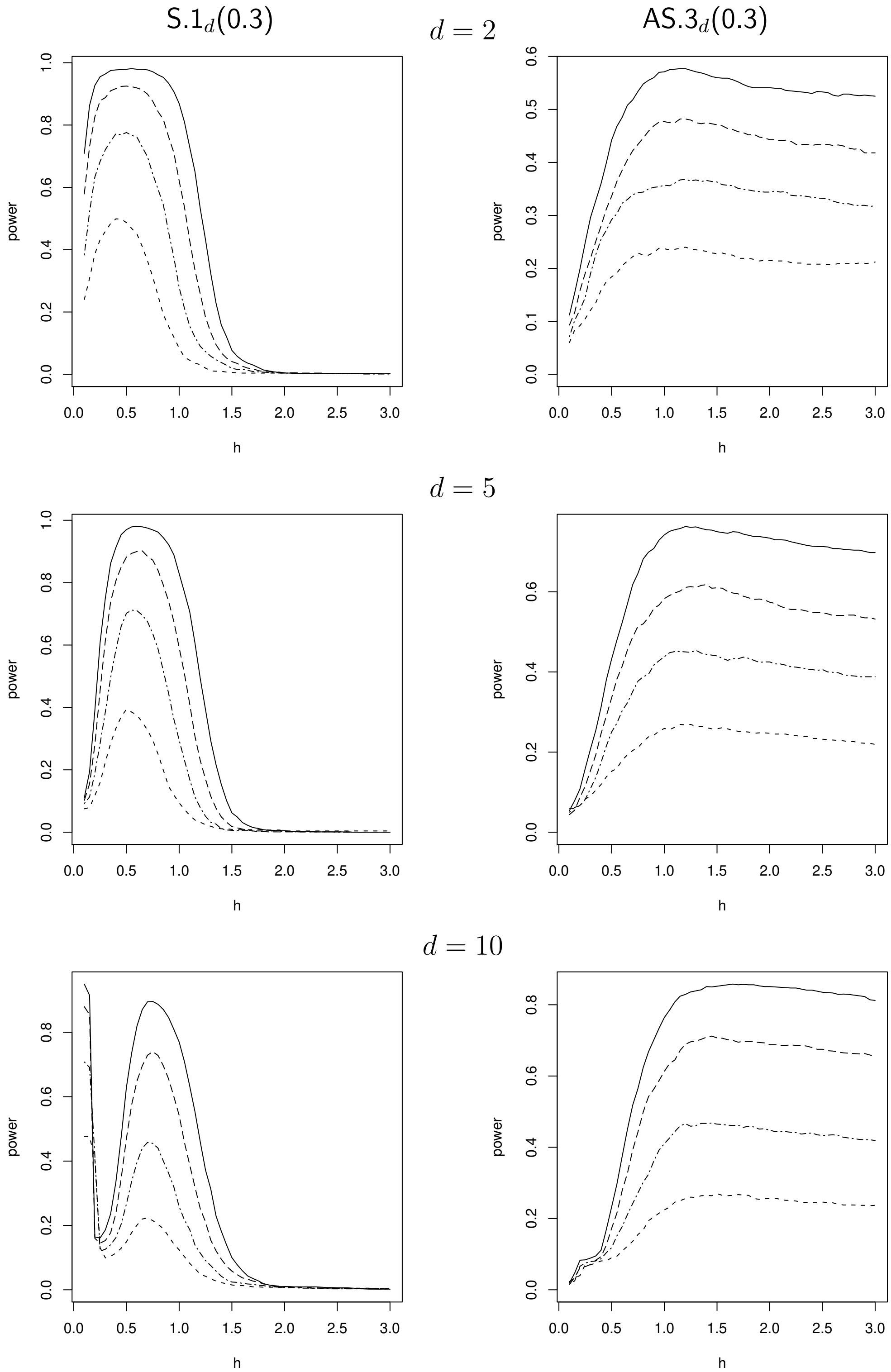

FIGURE 6. Empirical power as a function of h for $n = 100$ (solid line), $n = 80$ (large broken line), $n = 60$ (broken and dotted line), $n = 40$ (short broken line), and $d = 2, 5, 10$ at level 0.05.

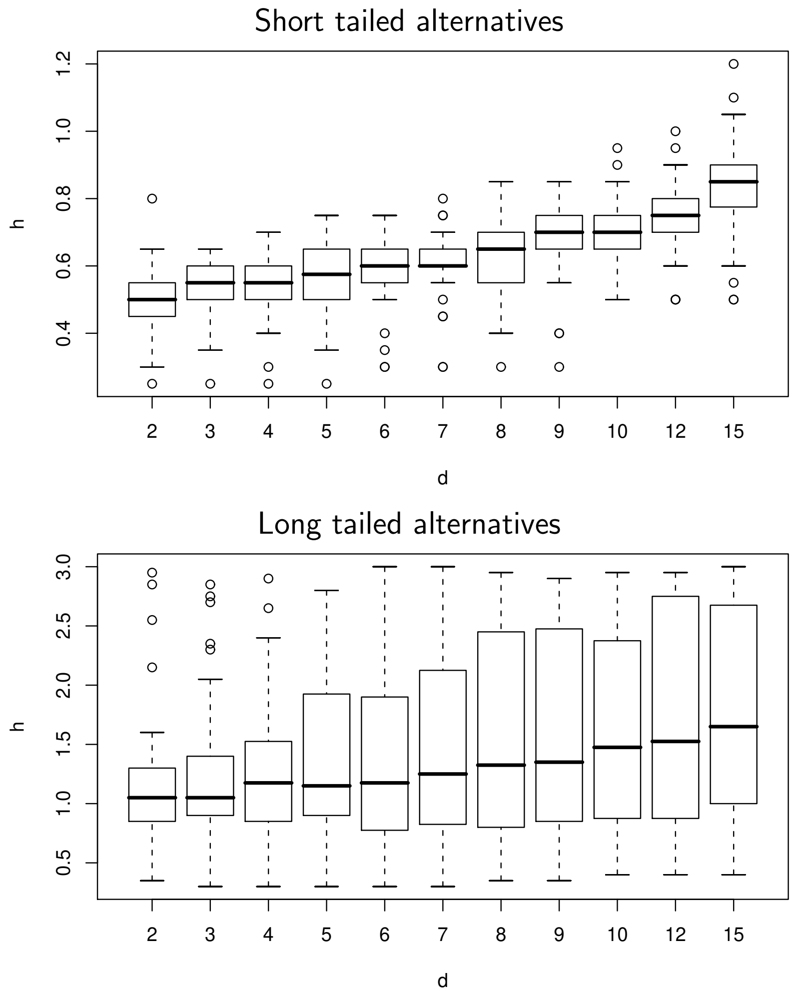

FIGURE 7. Distribution of $h_{f,n}$ as a function of d at level 0.05

power function of the test for an alternative f is an increasing function of $\sqrt{n}(\mathrm{E}_\phi \phi_{(1+2h^2)^{1/2}}(Y_1) - \mathrm{E}_{f^*}\phi_{(1+2h^2)^{1/2}}(Y_1))$ and therefore detect alternatives f satisfying $\int \phi_{(1+2h^2)^{1/2}}(u)\phi(u)du > \int \phi_{(1+2h^2)^{1/2}}(u)f^*(u)du$. This is in particular true for short tailed distributions.

In order to get an easy-to-use rule for choosing the bandwidth, we procede as in the univariate case. For data dimensions $d = 2, (1), 10, 12, 15$, for each one of the alternative distributions $\mathrm{S}.1_d(\rho)$,...,$\mathrm{AS}.4_d(\rho)$, where we take $\rho = 0, 0.3, 0.7$, and for n taking the values $n = 20, 40, 60, 80, 100$, we obtain the nearly optimal bandwidth $h_{f,n}$ defined in paragraph 3.4. However, when the empirical power present the behaviour shown by distribution $\mathrm{S}.1_{10}(0.3)$ in Figure 6, the small values of h that maximize the empirical power are not considered. In this case a large local maximum is taken for $h_{f,n}$.

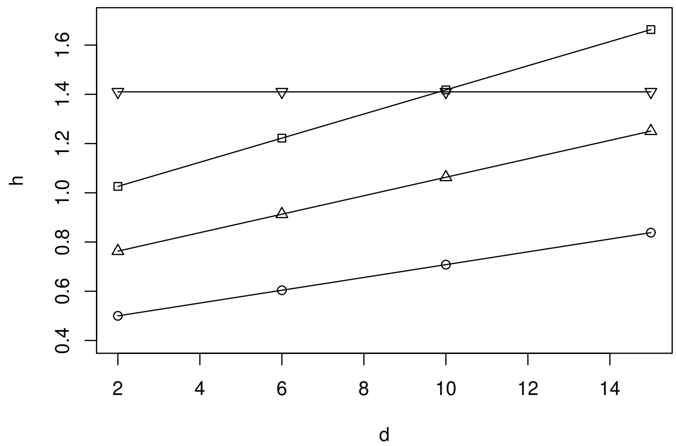

FIGURE 8. Considered choices of h as a function of d: $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\bigtriangledown)$

For each one of the considered data dimension d, the sample distribution of the bandwidths $h_{f,n}$ for the BHEP test at level 0.05 is described in Figures 7 for short tailed and long tailed alternatives, respectively. A linear regression of the corresponding sample medians over the data dimension, leads to the following relation that we will use as rule-of-thumb for the choice of h when the data dimension is d. For short tailed distributions we obtain

$$h_{\mathrm{S}} = 0.448 + 0.026d,$$

whereas for long tailed distributions the relation is

$$h_{\mathrm{L}} = 0.928 + 0.049d.$$

These rules-of-thumb can also be used for the BHEP test at levels 0.01 and 0.1, since the distributions of the bandwidths $h_{f,n}$ for these two levels are very similar to that one shown in Figure 7.

As in the univariate case, if we do not have relevante information about the tail of the alternative distribution, we propose to choose $h = \bar{h} := (h_{\mathrm{S}}+h_{\mathrm{L}})/2$. In Figure 8 these bandwidths are plotted as a function of d. Contrary to the usual choice $h_{\mathrm{HZ}} = 1.41$ proposed by Henze and Zirkler [23] that seems to be more suitable for long tailed alternatives, we expect that $\bar{h}$ could produce a omnibus test of normality showing a reasonable performance against a large set of alternative distributions.

4.2. Finite sample power analysis. A simulation was conducted to compare the power of the BHEP tests for the following choices of the smoothing

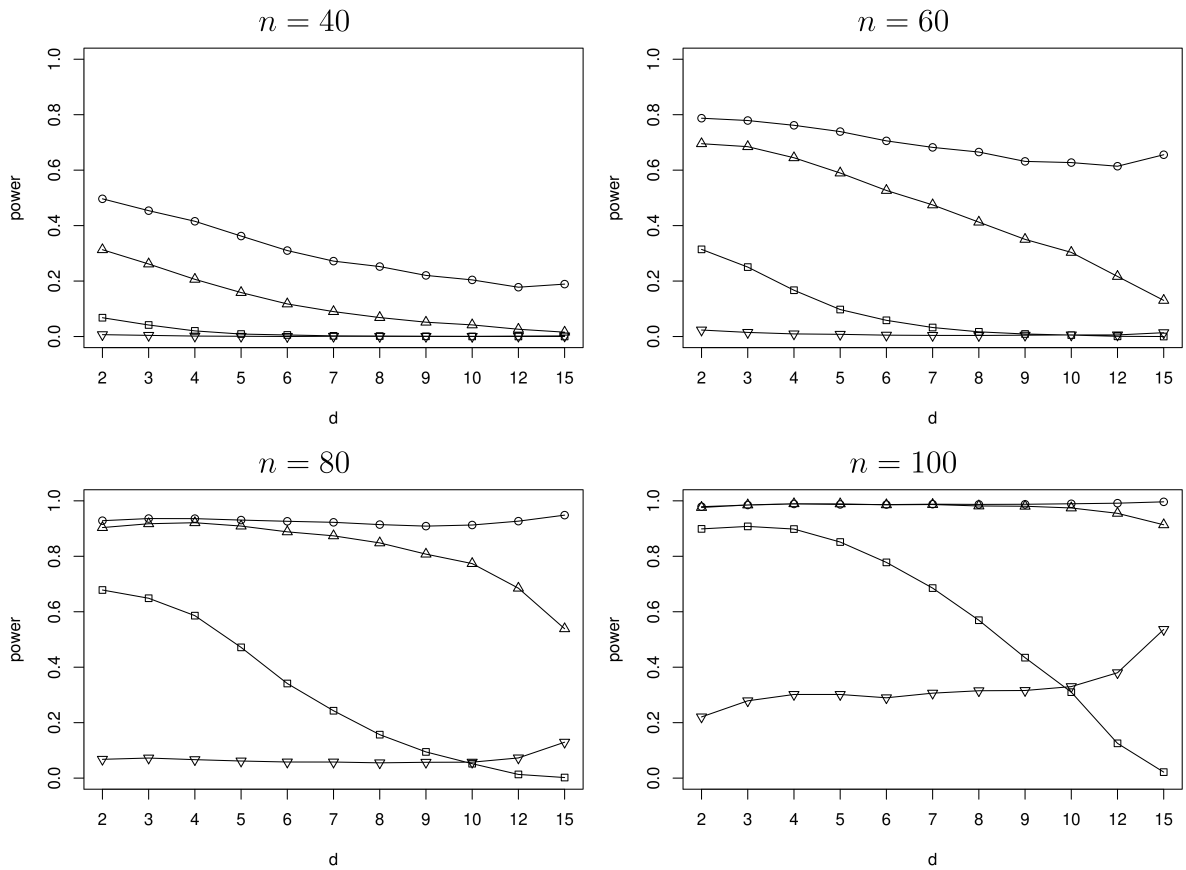

FIGURE 9. Empirical power for the multivariate uniform distribution from the Pearson Type II family for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\bigtriangledown)$

parameter: $h = h_{\mathrm{S}}, h_{\mathrm{L}}, \bar{h}, h_{\mathrm{HZ}}$. The set of alternative distributions considered include all the alternatives proposed by Mecklin and Mundfrom [28] and also other alternatives some of them previously considered in the simulation studies of Henze and Zirkler [23] and Romeu and Ozturk [32]. The Mecklin and Mundfrom's set of distributions, also considered in Farrel *al.* [19], include a large variety of alternatives which are described in detail in [28]. We used the algorithms described in Johnson [27] to generate all the multivariate distributions. The evaluation of the test statistic critical values was based on 10^5 repetitions under the null hypothesis and the power results are based on 10^4 samples of different sizes from the considered set of alternatives. The standard level of significance $\alpha = 0.05$ was used.

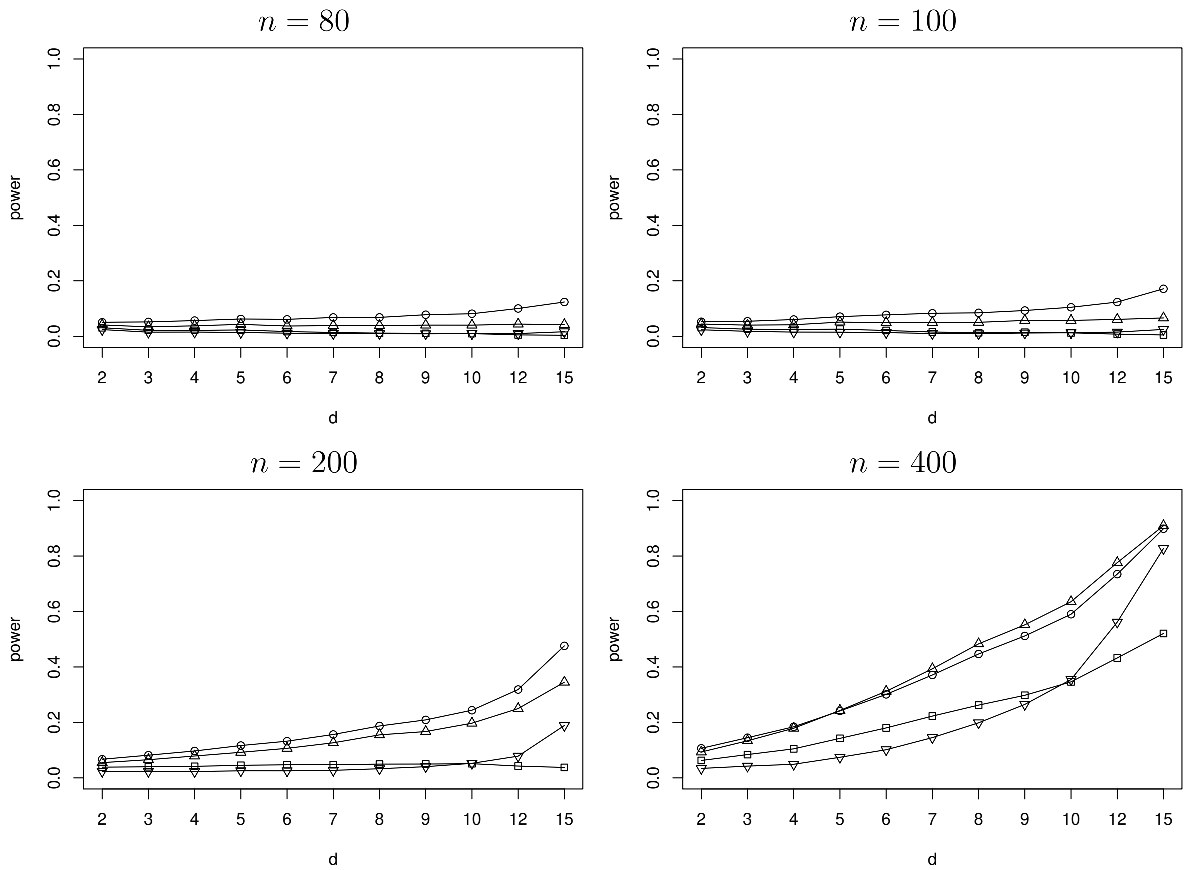

FIGURE 10. Empirical power for the Pearson Type II distribution with $m = 10$ for $h = h_{\rm S}(\bigcirc)$, $h = h_{\rm L}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\rm HZ}(\bigtriangledown)$

The pattern revealed by the tests in analysis was very clear: i) For short tailed distributions the best results were obtained by $h_{\rm S}$. The bandwidth $\bar{h}$ is better, and some times much better, than $h_{\rm HZ}$. For lower dimensions d the test based in $h_{\rm HZ}$ reveals a power performance against these alternatives. Figures 9 and 10 illustrate this situation showing the empirical powers for the Pearson Type II distributions with $m = 0$ and $m = 10$ (see [27] chapter 6). Contrary to the simulation results reported in Mecklin and Mundfrom [28] and Farrel *al.* [19] for the multivariate uniform distribution, the test based in $h_{\rm HZ}$ reveals a poor performance. This behaviour, which was also observed for all the considered members of the Pearson Type II family, agree with the simulation results reported in Henze and Zirkler [23]. Similarly to the univariate case, the previous behaviour was also observed for a high

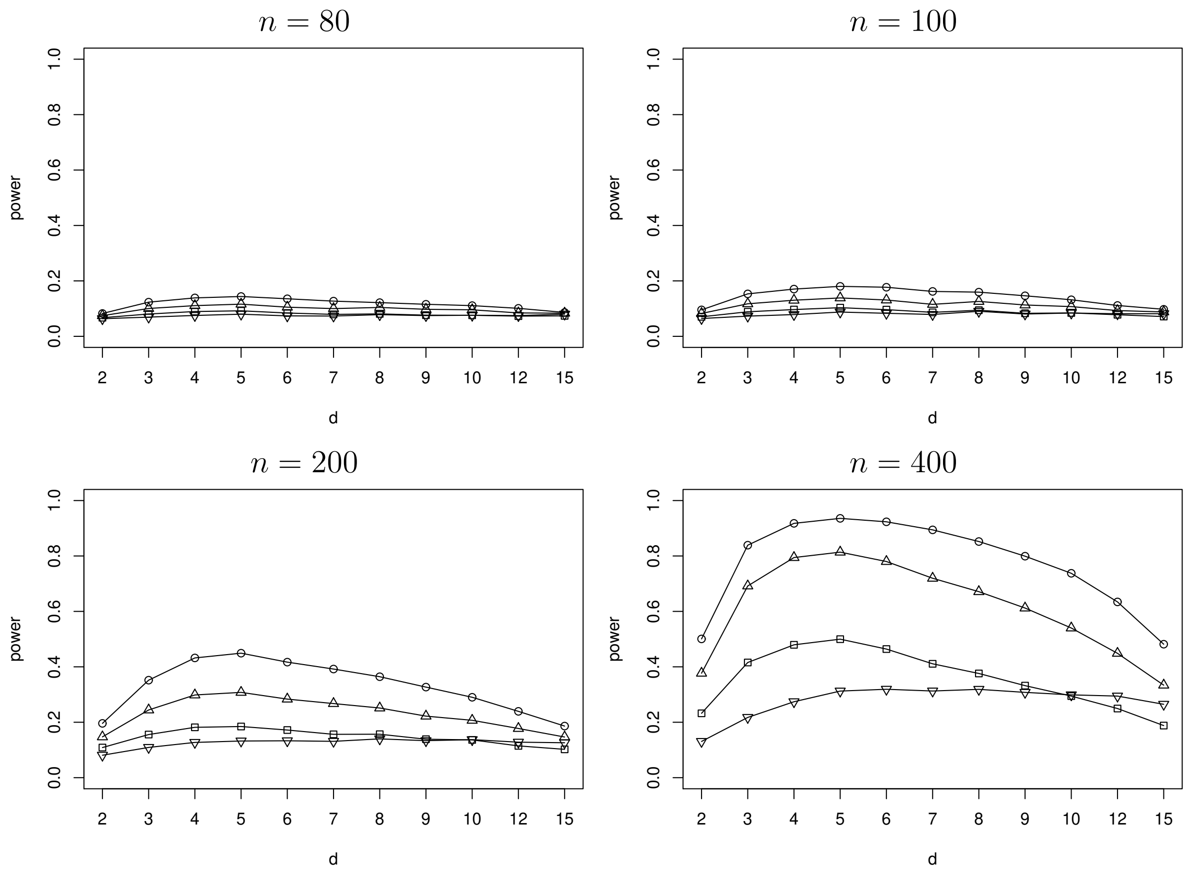

FIGURE 11. Empirical power for the high moment Khintchine distribution with GPE marginals for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\bigtriangledown)$

moment alternative like the Khintchine distribution with GEP marginals (see [27] chapter 8 and paragraph 2.4) as we can see in Figure 11. This non-normal distribution present an interesting departure from multivariate normality since the values of Mardia's skewness and kurtosis are equal to the one of the multivariate normal distribution. ii) For long tailed or moderately skewed alternatives the best results were obtained by the bandwidths h_{L} and h_{HZ}. However, for large values of d ($d > 10$) h_{L} is better than h_{HZ}. Although inferior to h_{HZ}, the bandwidth $\bar{h}$ reveled a reasonable performance against all the considered long tailed alternatives. Figures 12 and 13 present the power estimates for the multivariate Student distribution t_{10} from the Pearson Type VII family (see [27] chapter 6) and for the asymmetric multivariate Burr-Pareto-Logistic distribution with normal marginals with parameter $\alpha = 1$ (see [27] chapter 7). Finally, Figure 14 reports the power estimates for one

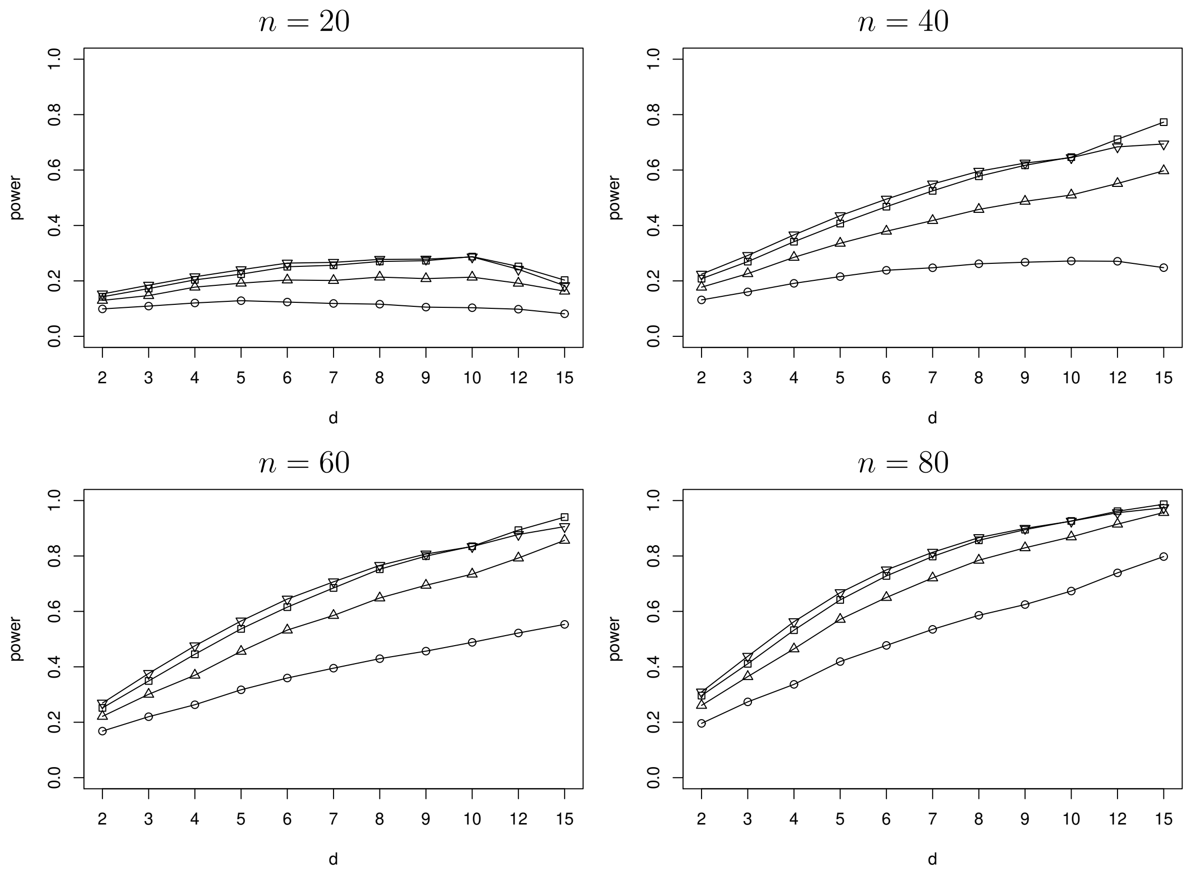

FIGURE 12. Empirical power for the multivariate Student distribution t_{10} for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\triangledown)$

of the fifteen different mixtures of two multivariate normals considered in Mecklin and Mundfrom [28] (normal mixture 13). This mixture reflects a mild contamination level and is skewed and leptokurtic.

4.3. Conclusion. The previous results show that the usual choice $h = h_{\mathrm{HZ}}$ gives the best results for long tailed or moderately skewed alternatives but it also produce very weak results for short tailed alternatives. In practice, if no relevant information about the tail of the alternatives is available, we strongly recommend the use of $\bar{h}$ that should replace, in a multivariate context, the usual choice of the bandwidth in the BHEP goodness-of-fit test. With this new bandwidth the BHEP test present a reasonable performance against all the considered alternatives. Moreover, for short tailed alternatives and long

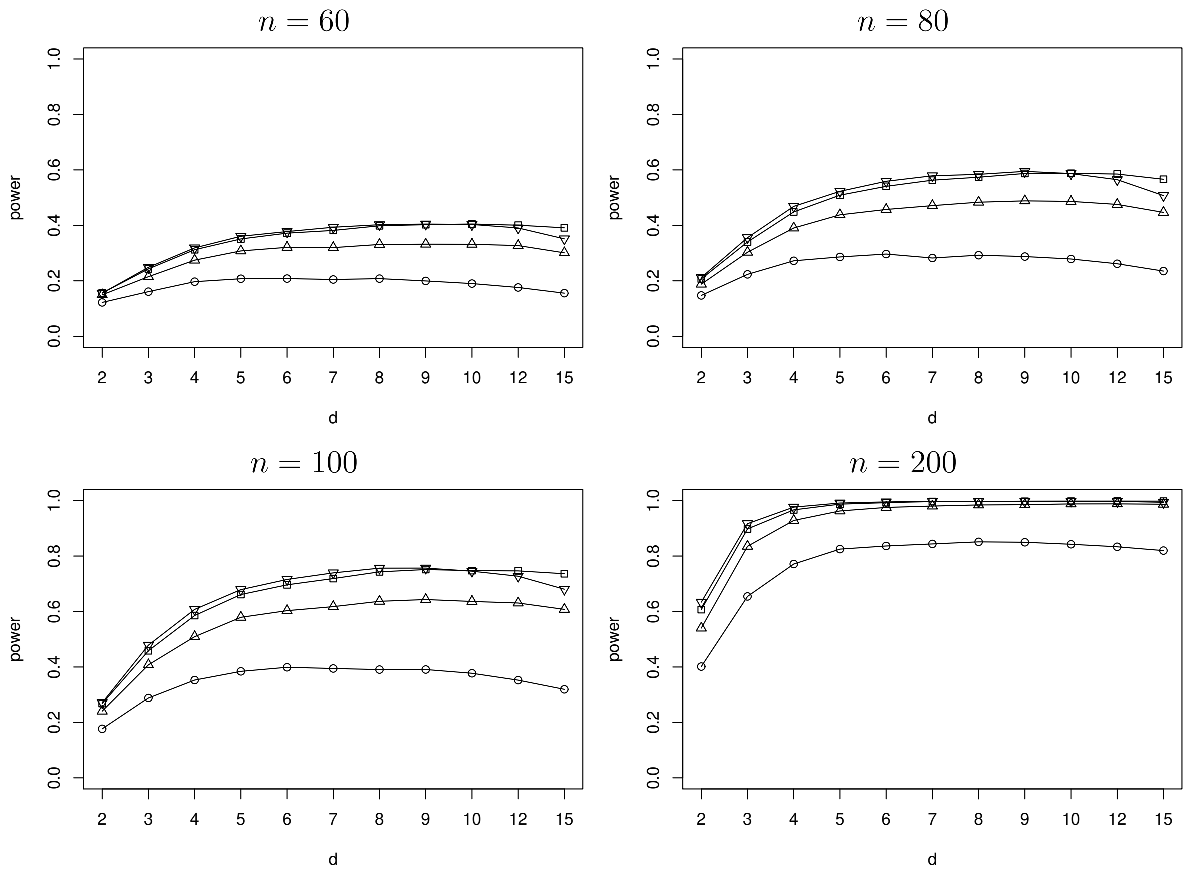

FIGURE 13. Empirical power for the multivariate Burr-Pareto-Logistic distribution with normal marginals with $\alpha = 1$ for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\bigtriangledown)$

tailed or moderately skewed alternatives the choices $h = h_{\mathrm{S}}$ and $h = h_{\mathrm{L}}$, respectively, are recommended.

5. Proof of Theorem 1

The following result follows straightforward from Lemma 2.4 of Gregory [21] and Bahadur's [5] results.

Lemma 1. *Let $X_1, X_2, \ldots$ a sequence of independent and identically distributed random variables whose distribution is determined by a parameter θ taking on values in a parametric set Θ. Let $T = (T_n)$ with $T_n = T_n(X_1, \ldots, X_n) \geq 0$, a sequence of test statistics to test $\theta \in \Theta_0$ against $\theta \in \Theta \setminus \Theta_0$, with $\Theta_0 \subset \Theta$, that satisfies the following conditions:*

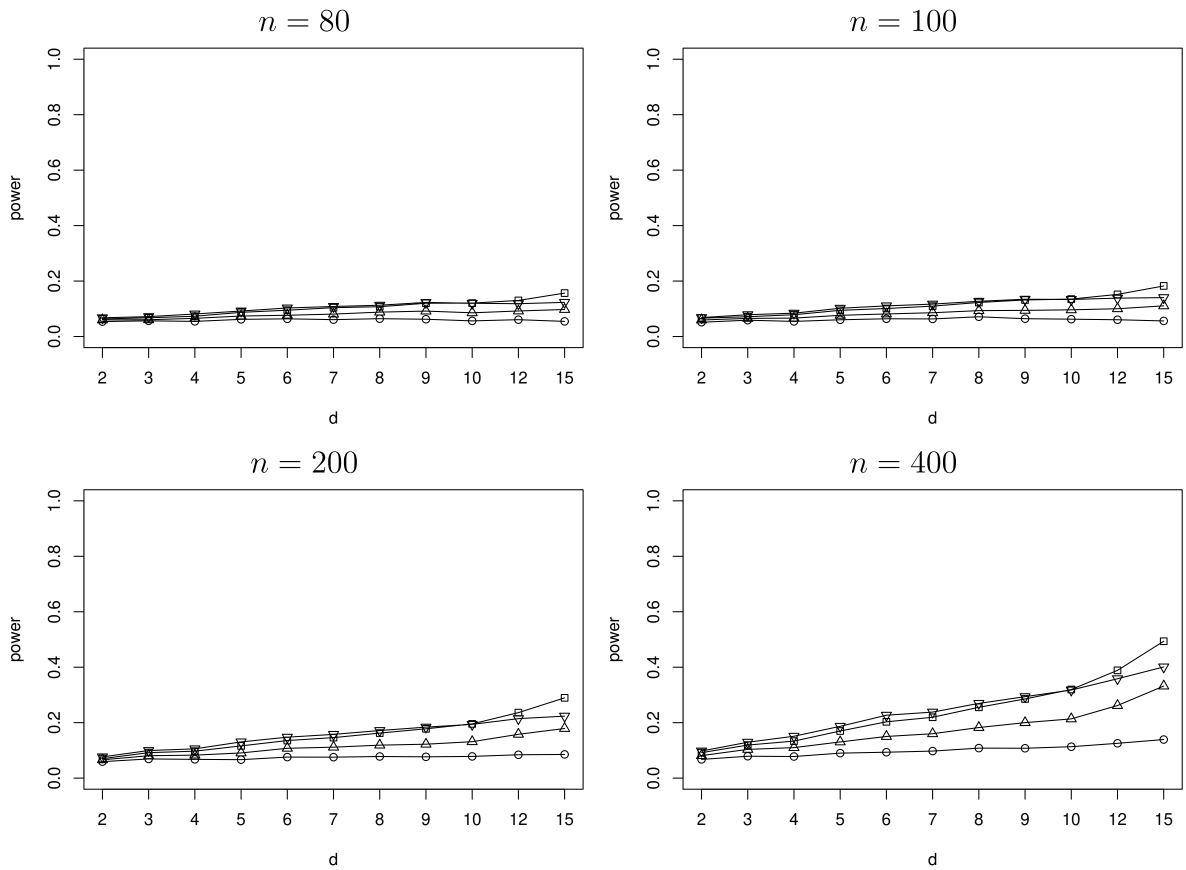

FIGURE 14. Empirical power for the mixture of normals distribution for $h = h_{\mathrm{S}}(\bigcirc)$, $h = h_{\mathrm{L}}(\square)$, $h = \bar{h}(\triangle)$ and $h = h_{\mathrm{HZ}}(\bigtriangledown)$

A. *For all* $\theta \in \Theta_0$ *we have*

$$T_n^2 \xrightarrow[n\to+\infty]{d} \mu + \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\mu \geq 0$, $\lambda_k \geq 0$, $k \in \mathbb{N}$, $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$, *and* (Z_k) *are iid standard normal random variables.*

B. *For all* $\theta \in \Theta \setminus \Theta_0$,

$$\frac{T_n^2}{n} \xrightarrow[n\to+\infty]{p} B(\theta),$$

for some real and positive function B *on* $\Theta \setminus \Theta_0$.

Then the Bahadur approximate slope of T *is given by*

$$C_T(\theta) = \lambda_1^{-1} B(\theta),$$

for all $\theta \in \Theta \setminus \Theta_0$.

Taking in Lemma 1, $\Theta = \mathcal{D}$, $\Theta_0 = \mathcal{D}_0$ and $T_n = \mathbf{I}_n^2(h)^{1/2}$, Theorem 1 follows easily from Theorems 1 and 2 of Tenreiro [35] since conditions **A**. and **B**. are satisfied with

$$\mu = \int \mathcal{Q}(u, u; h) d\Phi(u),$$

$$\lambda_k = \lambda_{k,h},$$

and

$$B(f) = (2\pi)^d \int |\hat{f}^*(t) - \hat{\phi}(t)|^2 |\hat{K}(ht)|^2 dt.$$

References

[1] Anderson, N.H., Hall, P., Titterington, D.M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* 50, 41–54.

[2] Arcones, M.A. (2006). On the Bahadur slope of the Lilliefors and the Cramér-von Mises tests of normality. In: IMS Lecture Notes – *High Dimensional Probability*, Evarist Giné, Vladimir Koltchinskii, Wenbo Li, Joel Zinn, Eds., 51, 196–206.

[3] Arcones, M.A., Wang, Y. (2006) Some new tests for normality based on U-processes. *Statis. Probab. Lett.* 76, 69-82.

[4] Arcones, M.A. (2007). Two tests for multivariate normality based on the characteristic function. *Math. Methods Statist.* 16, 177–201.

[5] Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* 31, 276–295.

[6] Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* 38, 303–324.

[7] Bahadur, R.R. (1971). *Some Limit Theorems in Statistics*. Philadelphia, SIAM.

[8] Baringhaus, L., Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35, 339–348.

[9] Baringhaus, L., Danschke, R., Henze, N. (1989). Recent and classical tests for normality. A comparative study. *Comm. Statist. Simulation Comput.* 18, 363–379.

[10] Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* 1, 1071–1095.

[11] Bowman, A.W., Foster, P.J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.* 88, 529–537.

[12] Coin, D. (2007). A goodness-of-fit test for normality based on polynomial regression. *Comput. Statist. Data Anal.* 52, 2185–2198.

[13] Csörgő, S. (1989). Consistency of some tests for multivariate normality. *Metrika* 36, 107–116.

[14] Durbin, J., Knott, M., Taylor, C.C. (1975). Components of Cramér-von Mises Statistics, II. *J. Roy. Statist. Soc. Ser. B* 37, 216–237.

[15] Epps, T.W., Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* 70, 723–726.

[16] Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* 10, 316–356.

[17] Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* 14, 604–621.

[18] Fang, H.-B., Fang, K.-T., Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* 82, 1–16.

[19] Farrel, P.J., Salibian-Barrera, M., Naczk, K. (2007). On tests for multivariate normality and associated simulation studies. *J. Stat. Comput. Simul.* 77, 1065–1080.

[20] Gouriéroux, C., Tenreiro, C. (2001). Local power properties of kernel based goodness of fit tests. *J. Multivariate Anal.* 78, 161–190.

[21] Gregory, G.G. (1980). On efficiency and optimality of quadratic tests. *Ann. Statist.* 8, 116–131.

[22] Hall, P. (1997). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

[23] Henze, N., Zirkler, B. (1990). A class of invariante consistent tests for multivariate normality. *Comm. Stat. Theory Methods* 19, 3595–3617.

[24] Henze, N. (1997). Extreme smoothing and testing for multivariate normality. *Statist. Probab. Lett.* 35, 203–213.

[25] Henze, N., Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *J. Multivariate Anal.* 62, 1–23.

[26] Johnson, M.E., Tietjen, G.L., Beckman, R.J. (1980). A new family of probability distributions with applications to Monte Carlo studies. *J. Amer. Statist. Assoc.* 75, 276–279.

[27] Johnson, M.E. (1987). *Multivariate Statistical Simulation.* Wiley, New York.

[28] Mecklin, C.J., Mundfrom, D.J. (2005). A Monte Carlo comparison of Type I and Type II error rates of tests of multivariate normality. *J. Stat. Comput. Simul.* 75, 93–107.

[29] Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press.

[30] R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

[31] Ramberg, J.S., Schmeiser, B.W. (1974). An approximate method for generating asymmetric random variables. *Commun. ACM* 17, 78–82.

[32] Romeu, J.L., Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *J. Multivariate Anal.* 46, 309–334.

[33] Stephens, M.A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.* 4, 357–369.

[34] Tenreiro, C. (2005). On the role played by the fixed bandwidth in the Bickel-Rosenblatt goodness-of-fit test. *SORT* 29, 201–216.

[35] Tenreiro, C. (2007). On the asymptotic behaviour of location-scale invariant Bickel-Rosenblatt tests. *J. Statist. Plann. Inference* 137, 103–116.

[36] Yazici, B., Yolacan, S. (2007). A comparison of various tests of normality *J. Stat. Comput. Simul.* 77, 175–183.

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3001–454 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt
URL: http://www.mat.uc.pt/∼tenreiro/